UDC [621.372.54::681.325.5]:681.513.673
Electronics and Communications in Japan, Vol. 67-A, No. 5, 1984
Translated from Denshi Tsushin Gakkai Ronbunshi, Vol. 67-A, No. 2, February 1984, pp. 126-132

p. 19-27 = (9)
PD: 05-84

An Adaptive Filtering Algorithm Using an Orthogonal Projection to an Affine Subspace and Its Properties

Kazuhiko Ozeki and Tetsuo Umeda, Regular Members

NHK Technical Research Laboratories, Tokyo, Japan 167

SUMMARY

The LMS algorithm and learning identification, which presently are typical adaptive algorithms, have a problem in that the speed of convergence may decrease greatly depending on the property of the input signal. To avoid this problem, this paper presents a geometrical discussion as to the origin of that defect, and proposes a new adaptive algorithm based on the result of the investigation. Comparing the convergence speeds of the proposed algorithm and the learning identification by numerical experiment by computer, great improvement was verified. The algorithm is extended to a group of algorithms which includes the original algorithm and the learning identification, which are called APA (affine projection algorithm). It is shown that APA has some desirable properties, such as, the coefficient vector approaches the true value monotonically and the convergence speed is independent of the amplitude of the input signal. Clear conclusions are also obtained for the problem as to what noise is included in the output signal when an external disturbance is impressed or the degree of the adaptive filter is not sufficient.

1. Introduction

The filter with a kind of learning function, in which the input signal and the output signal to produce (desirable output) are specified and the coefficients of the filter are modified successively so that the output approaches the desired output, is called adaptive filter. It has been applied to many problems such as automatic equalizer, echo canceller, and noise-elimination device.

The most important problem in the adaptive filter is the algorithm of how to modify the coefficients successively. A number of studies has been made on the algorithm, the most well-known being the LMS algorithm [1]. The algorithm has the feature in that it is simple, requires less computation time and is easy to implement on hardware. The adaptive filter using this algorithm is already on the market. On the other hand, it has a problem in that the convergence speed cannot be made very high, and it is not suited to application requiring a fast convergence. It has another problem in that the convergence speed depends greatly on the property of the input signal.

Another algorithm similar to LMS algorithm is the learning identification [2]. This algorithm can be regarded as an improvement of the LMS algorithm. Although the computational complexity increases, it has many desirable features in that the convergence speed is higher, and is independent of the amplitude of the input signal. An attempt has been made to apply the algorithm to the echo-canceller [3]. However, the learning identification has a problem common to LMS algorithm in that the convergence speed is degraded sometimes depending on the property of the input signal.

This paper is an attempt to solve the problem in those typical adaptive algorithms up to the present where the convergence speed may be degraded. By analyzing the algorithm from a geometrical viewpoint, a new adaptive algorithm is derived. Its effectiveness is verified and its characteristics are discussed.

Section 2 outlines the learning identification and LMS algorithm, and describes their problems. Section 3 derives a new adaptive algorithm based on the geometrical considerations. The effectiveness of the new algorithm is verified by a numerical experiment by computer.

ISSN0424-8368/83/0005-0019$7.50/0

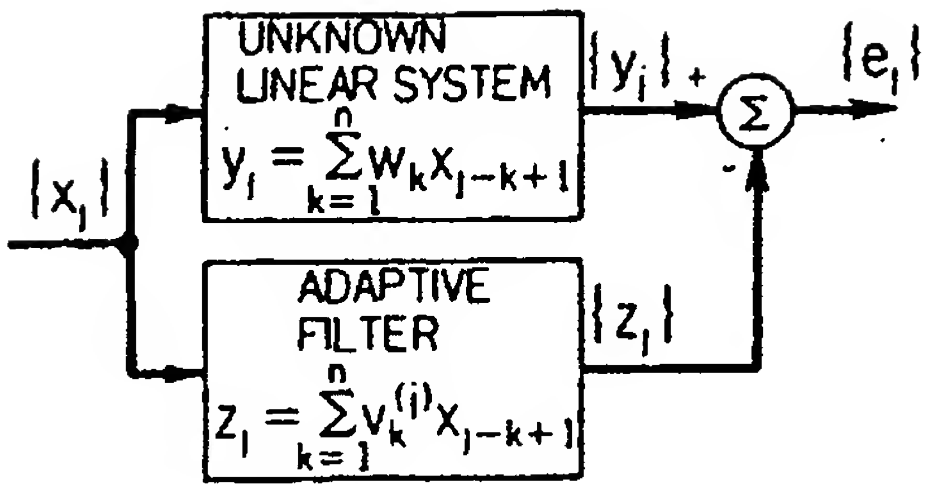

Fig. 1. Linear system identification by an adaptive filter.

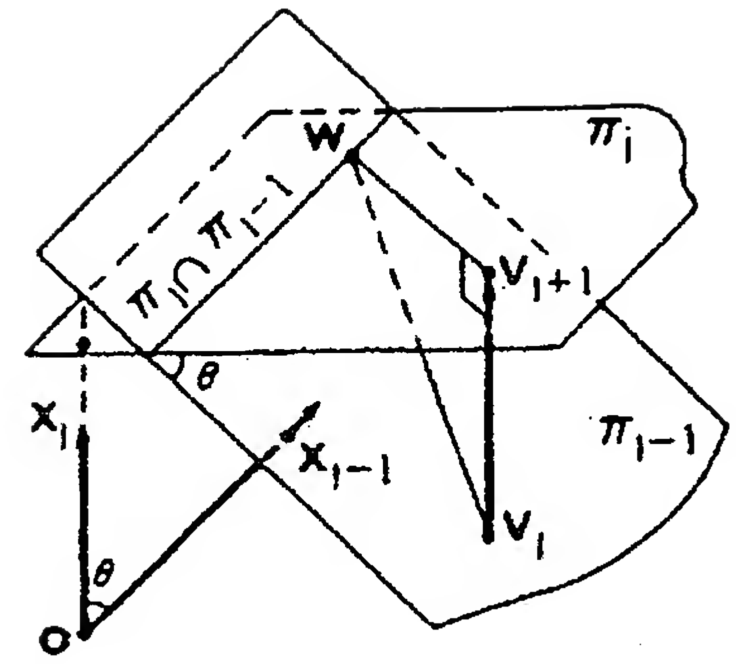

Fig. 2. Geometrical illustration of the learning method ($\mu = 1$, $n = 3$).

In Sect. 4, the algorithm derived in solution 3 is examined from the viewpoint of orthogonal projection to affine subspace. Then the algorithm is extended to a more general algorithm. Several practically important properties are discussed, including the behavior under external disturbance.

2. Past Algorithm

The algorithm proposed in this paper can be considered as an extension of the learning identification. The learning identification is outlined first, and then the LMS algorithm is reviewed within that framework. The problems of those algorithms are described from a geometrical viewpoint, which gives way to the new algorithm proposed in this paper.

2.1 Formulation of the problem and notations

The adaptive filter can be formulated as a problem of identifying a linear system. Consider, as in Fig. 1, an unknown system to be identified, which gives the output $\{y_j\}$ determined by

$$y_j = \sum_{k=1}^{n} w_k x_{j-k+1}$$

for input signal $\{x_j\}$. In the above, w_1, w_2, $\cdots$, w_n are unknown constants. The signal $\cdots$, x_{-1}, x_0, x_1, x_2, $\cdots$ is represented by $\{x_j\}$ and $\cdots$, y_{-1}, y_0, y_1, y_2, $\cdots$ is represented by $\{y_j\}$.

Then consider another linear system in which the output $\{z_j\}$ for the same input signal $\{x_j\}$ is determined by

$$z_j = \sum_{k=1}^{n} v_k^{(j)} x_{j-k+1}$$

It is called adaptive filter. The adaptive algorithm is used to modify the coefficient vector of the adaptive filter at time j

$$\boldsymbol{v}_j \triangleq (v_1^{(j)}, v_2^{(j)}, \cdots, v_n^{(j)})^t$$

in the form

$$\boldsymbol{v}_{j+1} = \boldsymbol{v}_j + \Delta \boldsymbol{v}_j$$

successively, to make it approach the coefficient vector of the system to be identified:

$$\boldsymbol{w} \triangleq (w_1, w_2, \cdots, w_n)^t$$

The vector $\Delta \boldsymbol{v}_j$ in the above expression is restricted to a function of the input and output values up to time j:

$$\Delta \boldsymbol{v}_j = f(x_j, x_{j-1}, \cdots, x_{j-m}, y_j, y_{j-1}, \cdots, y_{j-m}, z_j, z_{j-1}, \cdots, z_{j-m})$$

By the choice of f, various kinds of algorithms can be obtained.

The following notations are used in addition to those already used:

(i) $\boldsymbol{x}_j \triangleq (x_j, x_{j-1}, \cdots, x_{j-n+1})^t$.

(ii) For $\boldsymbol{a} = (a_1, \cdots, a_n)^t$, and $\boldsymbol{b} = (b_1, \cdots, b_n)^t$

$$\langle \boldsymbol{a}, \boldsymbol{b} \rangle \triangleq \sum_{k=1}^{n} a_k b_k$$

(iii) $\| \boldsymbol{a} \| \triangleq \sqrt{\langle \boldsymbol{a}, \boldsymbol{a} \rangle}$

(iv) $\Pi_j \triangleq \{ \boldsymbol{v} ; \boldsymbol{v} \in R^n, \langle \boldsymbol{v}, \boldsymbol{x}_j \rangle = y_j \}$

The set Π_j defined in (iv) is the set of all coefficient vectors which give the output equal to y_j for the input vector $\boldsymbol{x}_j$, and forms a hyperplane in the n-dimensional Euclidean space.

2.2 Learning identification [2]

In the learning identification, the coefficient vector is modified as follows.

1° Setting of initial value: v_0 = arbitrary value.

2° Iteration:

2.1° $z_j = \langle v_j, x_j \rangle$

2.2° $e_j = y_j - z_j$

2.3° $\Delta v_j = \dfrac{e_j}{\| x_j \|^2} x_j$

2.4° $v_{j+1} = v_j + \mu \Delta v_j$

The constant μ is called the relaxation constant. As is shown in Fig. 2, when $\mu = 1$, v_{j+1} is the end of the vertical line from v_j to Π_j; $\|v_{j+1} - w\| \geq \|v_j - w\|$ holds when $\mu \leq 0$ or $\mu \geq 2$. Consequently, $0 < \mu < 2$ must be the case for the coefficient vector to converge to w. Then, $\|v_{j+1} - w\| \leq \|v_j - w\|$ and the convergence is monotonic. When $\{x_j\}$ is multiplied by a constant, Δv_j and consequently the convergence speed do not change. This is another desirable property of the learning identification.

2.3 LMS algorithm [1]

Step 2.3° of the learning identification is modified as

$$\Delta v_j = e_j x_j$$

which results in the LMS algorithm. The direction of coefficient modification is the same as that of the learning identification, and v_{j+1} is a point on the vertical line from v_j to Π_j.

In the past, LMS algorithm is understood as an approximation to the steepest descent method, but the geometrical viewpoint is better to understand its behavior. In LMS algorithm, the monotonic property of the convergence of the coefficient vector is not ensured, and the convergence speed depends on the amplitude of $\{x_j\}$. From this point, the learning identification is more desirable than the LMS algorithm.

2.4 Problems in learning identification and LMS algorithm

For simplicity, the learning identification with $\mu = 1$ is considered. As is seen from Fig. 2, the convergence speed of the coefficient vector depends greatly on the angle between Π_j and Π_{j-1}. In other words, when the angle between Π_j and Π_{j-1} approaches 0 or Π,

$$\frac{\| v_{j+1} - w \|}{\| v_j - w \|} \rightarrow 1$$

and the convergence speed is decreased. Let the angle between Π_j and Π_{j-1} be θ, which is also the angle between x_j and x_{j-1}:

$$\cos \theta = \frac{\langle x_j, x_{j-1} \rangle}{\| x_j \| \cdot \| x_{j-1} \|}$$

The right-hand side of the above equation is nothing but the first-order sample autocorrelation function of the signal $\{x_j\}$. Consequently, the convergence speed decreases as the first-order autocorrelation function of the signal approaches 1 in absolute value. The situation is the same for the case of $\mu \neq 1$ and in LMS algorithm.

This phenomenon arises because the direction of coefficient modification is restricted to that of x_j. To improve the situation, the direction of the coefficient modification should be reconsidered.

3. New Adaptive Algorithm and Its Convergence Speed

3.1 Construction of algorithm

As is seen in Fig. 2, to keep the convergence speed constant, independently of the angle between x_j and x_{j-1}, the vertical line should be drawn from v_j to $\Pi_j \cap \Pi_{j-1}$, not to Π_j. Let the end of the vertical line be v_{j+1} and introducing the relaxation constant μ in the same way as in the learning identification, the following algorithm can be constructed (Fig. 3):

1° Setting of initial value: v_0 = arbitrary value.

2° Iteration: $v_{j+1} = v_j + \mu (v_{j+1} - v_j)$. When $\mu = 1$, the iteration of this algorithm can be written as follows:

2.1° $\tilde{x}_{j-1} = \dfrac{\langle x_{j-1}, x_j \rangle}{\| x_{j-1} \|^2} x_{j-1}$

2.2° $u_j = x_j - \tilde{x}_{j-1}$

2.3° $z_j = \langle v_j, x_j \rangle$

2.4° $e_j = y_j - z_j$

2.5° $\Delta v_j = \dfrac{e_j}{\langle u_j, x_j \rangle} u_j$

2.6° $v_{j+1} = v_j + \Delta v_j$

The properties of this algorithm and the computation for $\mu \neq 1$ are discussed as the general theory in the next section.

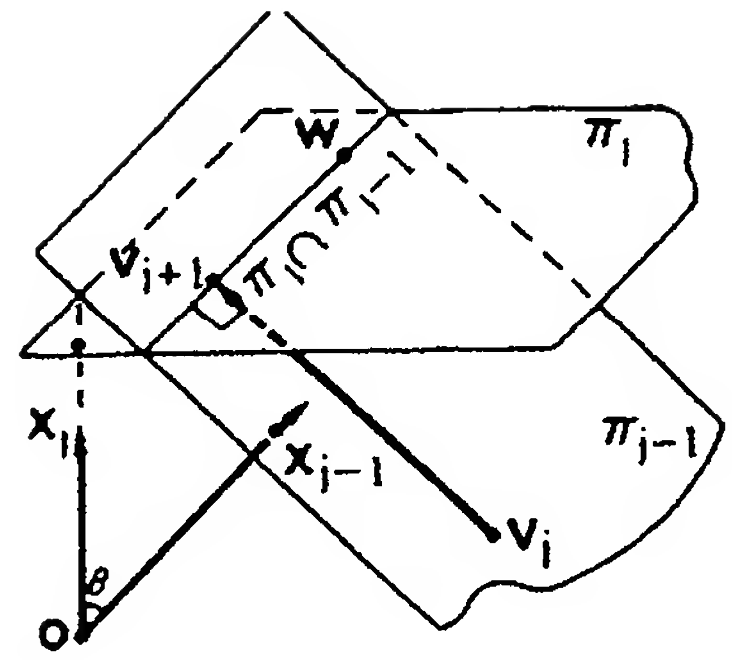

Fig. 3. Geometrical illustration of the new algorithm (μ = 1, n = 3).

3.2 Convergence speed

To compare the convergence speeds of the learning identification and the method proposed in this paper, a numerical experiment was performed by a computer. As is shown in Fig. 4, a colored noise, which is obtained by the normally distributed random numbers through a first-order recursive filter, is used as the signal $\{x_j\}$. The autocorrelation function of this signal is the filter coefficient α itself, and the convergence speed was examined by varying the value of α. Nearly the same coefficient vector is used for the system to be measured as is used in the numerical experiment in [2].

In the experiment, the following number of steps $j(\varepsilon)$ is determined, instead of the convergence speed, which is defined as follows:

$$j(\varepsilon)=\min\{j;\|v_j-w\|/\|w\|\leq\varepsilon\},\ v_0=\mathbf{0}$$

Figures 5(a) and (b) show the relation between ε and $j(\varepsilon)$ for various values of α for the learning identification and the new algorithm both with μ = 1. The order of the filter is set as 16. By this comparison, the following observations are made.

(i) When α = 0, there is no great difference, although the new algorithm gives a slightly faster convergence. As is seen from Figs. 2 and 3, if Π_j and Π_{j-1} are always orthogonal, the learning identification and the new algorithm are the same. When α = 0, the sitaution is close to this.

(ii) When α approaches 1, $j(\varepsilon)$ in learning identification rapidly increases, while it does not change much in the new algorithm. This is anticipated from the geometrical interpretation of the algorithm. When α = 0.99, the convergence speed of the new algorithm is more than 10 times the learning identification.

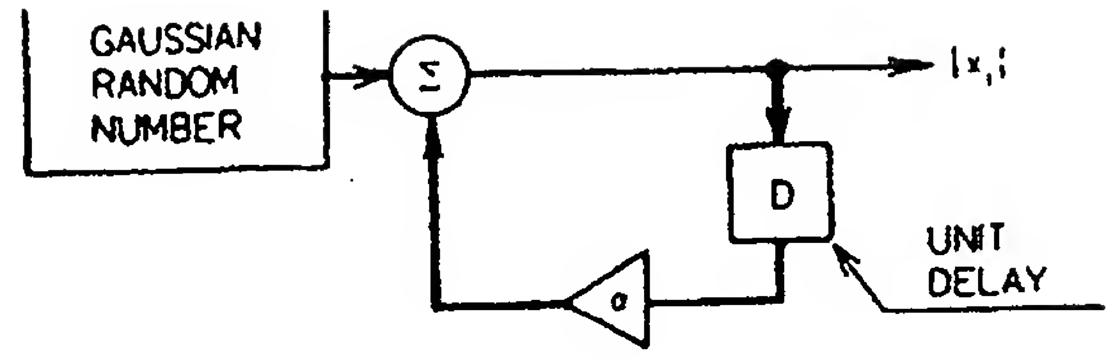

Fig. 4. A colored gaussian noise generated by a 1st-order recursive filter.

4. Extension of Algorithm [4, 5, 6]

4.1 APA (affine projection algorithm)

Π_j or $\Pi_j \cap \Pi_{j-1}$ used so far, does not necessarily contain the origin of $\mathbf{R}^n$. Consequently, it is not necessarily a subspace of $\mathbf{R}^n$ as a vector space, but an affine subspace. Let Π be an affine subspace of $\mathbf{R}^n$. The mapping, which is an orthogonal projection of $\mathbf{R}^n$ on Π is written as $\mathbf{P}_\Pi$. Using this notation, the learning identification for μ = 1 is written as the coefficient modification algorithm by

$$v_{j+1}=P_{\Pi_j}(v_j)$$

and the proposed algorithm is that by

$$v_{j+1}=P_{\Pi_j\cap\Pi_{j-1}}(v_j)$$

From such a viewpoint, these algorithms can easily be extended. An algorithm that performs the modification by

$$v_{j+1}=P_{\Pi_j\cap\Pi_{j-1}\cap\cdots\cap\Pi_{j-(p-1)}}(v_j) \qquad (1)$$

is considered. The vector $\mathbf{v}_{j+1}$ in Eq. (1) is the solution of the system of equations with $\mathbf{v}$ as unknowns:

$$\left.\begin{array}{l}\langle x_j,\ v\rangle=y_j\\ \langle x_{j-1},\ v\rangle=y_{j-1}\\ \quad\vdots\\ \langle x_{j-(p-1)},\ v\rangle=y_{j-(p-1)}\end{array}\right| \qquad (2)$$

which minimizes $\|\mathbf{v}-\mathbf{v}_j\|$. Letting the coefficient matrix of the left-hand side of Eq. (2) be

$$X_j=(x_j,\ x_{j-1},\cdots,\ x_{j-(p-1)})^t,$$

and the constant in the right-hand side be

$$y_j=(y_j,\ y_{j-1},\cdots,\ y_{j-(p-1)})^t$$

and letting $\mathbf{X}_j^+$ be the Moore-Penrose generalized inverse of $\mathbf{X}_j$, it can be written as [7]

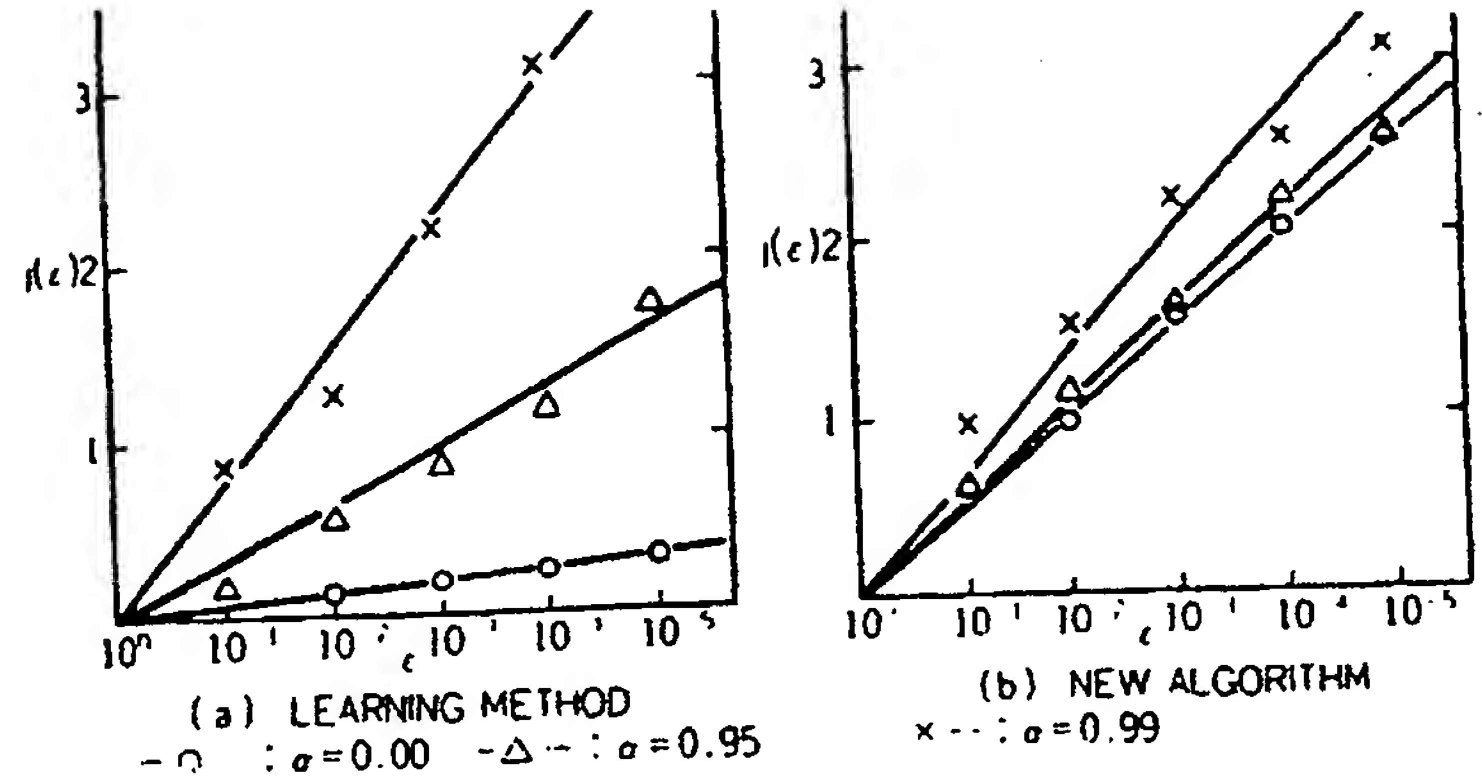

Fig. 5. Comparison of the convergence time between the learning method and the new algorithm.

$$v_{j+1} = X_j^+ y_j + (I - X_j^+ X_j) v_j \qquad I\, X_j^+$$
$$= v_j + X_j^+ (y_j - X_j v_j) \tag{3}$$

where I is the unit matrix.

Based on above equation and introducing the relaxation coefficient μ, the following adaptive algorithm is considered:

1° Setting of initial value: v_0 = arbitrary value.

2° Iteration:

2.1° $\Delta v_j = X_j^+ (y_j - X_j v_j)$

2.2° $v_{j+1} = v_j + \mu \Delta v_j$

In this paper, this algorithm is called APA (affine projection algorithm) and p is called its order. According to this definition, the learning identification is the first-order APA, and the algorithm in the preceding section is the second-order APA.

4.2 Fundamental properties of APA

This section describes three fundamental properties of APA.

Property 1. If $0 < \mu < 2$, $\|v_{j+1} - w\| \leq \|v_j - w\|$. If $\mu \leq 0$ or $\mu \geq 2$, $\|v_{j+1} - w\| \geq \|v_j - w\|$.

Proof. Let $\tilde{v}_{j+1} = v_j + \Delta v_j$. Then

$$\tilde{v}_{j+1} = P_{\Pi_j \cap \Pi_{j-1} \cap \cdots \cap \Pi_{j-(p-1)}} (v_j)$$

Since $w \in \Pi_j \cap \Pi_{j-1} \cap \cdots \cap \Pi_{j-(p-1)}$, $\tilde{v}_{j+1} - w$ and Δv_j are orthogonal to each other. Consequently, by Pythagoras' theorem,

$$\|v_j - w\|^2 = \|\Delta v_j\|^2 + \|\tilde{v}_{j+1} - w\|^2,$$
$$\|v_{j+1} - w\|^2 = \|\tilde{v}_{j+1} - v_{j+1}\|^2 + \|\tilde{v}_{j+1} - w\|^2$$
$$= (1-\mu)^2 \|\Delta v_j\|^2 + \|\tilde{v}_{j+1} - w\|^2 \tag{4}$$

Consequently,

$$\|v_j - w\|^2 - \|v_{j+1} - w\|^2 = \mu(2-\mu)\|\Delta v_j\|^2$$

Thus, the result is obtained. (End of Proof.)

From the above property, it is seen that in order for the coefficient vector in APA to converge to w, $0 < \mu < 2$ is necessary It is seen also that if μ is in this range, the coefficient vector never goes away from w, i.e., the convergence is monotonic. It i not necessarily true that $0 < \mu < 2$ is the s ficient condition for the convergence.

Property 2. Let $0 < \mu < 2$ and $p > q$. Let the coefficient vector v_j be modified fc once by pth and qth order APA, and the resul ing coefficient vectors be $v_{j+1}^{(p)}$ and v_{j+1} respectively. Then,

$$\|v_{j+1}^{(p)} - w\| \leq \|v_{j+1}^{(q)} - w\|$$

Proof. Let $\mu = 1$, and the coefficien vectors obtained by modifying v_j as above b $\tilde{v}_{j+1}^{(p)}$ and $\tilde{v}_{j+1}^{(q)}$, respectively. Then by the same reasoning as in Eq. (4),

$$\|v_{j+1}^{(p)} - w\|^2 = \|\tilde{v}_{j+1}^{(p)} - v_{j+1}\|^2 + \|\tilde{v}_{j+1}^{(p)} - w\|^2$$
$$= \mu(2-\mu)\|\tilde{v}_{j+1}^{(p)} - w\|^2$$
$$+ (1-\mu)^2 \|v_j - w\|^2$$

Similar relation applies to $\|v_{j+1}^{(q)} - w|$ Consequently,

$$\|v_{j+1}^{(p)} - w\|^2 - \|v_{j+1}^{(q)} - w\|^2$$
$$= \mu(2-\mu)\{\|\tilde{v}_{j+1}^{(p)} - w\|^2 - \|\tilde{v}_{j+1}^{(q)} - w\|^2\} \qquad (5)$$

On the other hand, by the property of the orthogonal projection,

$$P_{\Pi} \circ P_{\Pi'} = P_{\Pi'} \circ P_{\Pi} = P_{\Pi}$$

holds in general for affine subspace $\Pi \supset \Pi'$, where $\circ$ indicates the composition of mappings. Since $p > q$ is assumed,

$$\Pi_j \cap \Pi_{j-1} \cap \cdots \cap \Pi_{j-(p-1)} \subset \Pi_j \cap \Pi_{j-1} \cap \cdots \cap \Pi_{j-(q-1)}$$

Consequently,

$$\begin{aligned}\tilde{v}_{j+1}^{(p)} &= P_{\Pi_j \cap \Pi_{j-1} \cap \cdots \cap \Pi_{j-(p-1)}}(v_j) \\ &= P_{\Pi_j \cap \Pi_{j-1} \cap \cdots \cap \Pi_{j-(p-1)}} \\ &\quad \circ P_{\Pi_j \cap \Pi_{j-1} \cap \cdots \cap \Pi_{j-(q-1)}}(v_j) \\ &= P_{\Pi_j \cap \Pi_{j-1} \cap \cdots \cap \Pi_{j-(p-1)}}(\tilde{v}_{j+1}^{(q)})\end{aligned}$$

It is seen from this that $\tilde{v}_{j+1}^{(p)} - w$ and $\tilde{v}_{j+1}^{(p)} - \tilde{v}_{j+1}^{(q)}$ are orthogonal, and

$$\begin{aligned}\|\tilde{v}_{j+1}^{(q)} - w\|^2 &= \|\tilde{v}_{j+1}^{(q)} - \tilde{v}_{j+1}^{(p)}\|^2 + \|\tilde{v}_{j+1}^{(p)} - w\|^2 \\ &\geqq \|\tilde{v}_{j+1}^{(p)} - w\|^2\end{aligned}$$

Using this inequality, Eq. (5) and $0 < \mu < 2$, the result is obtained. (End of Proof.)

It is anticipated from Property 2 that the convergence speed may be increased by increasing the order.

Property 3. Let $\{\tilde{x}_j\}$ be the signal obtained by multiplying the amplitude of the input signal $\{x_j\}$ by a $(a \neq 0)$, i.e., $\tilde{x}_j = a\, x_j$. Starting from the initial value v_0, let the coefficient vector obtained by using the input signal $\{x_j\}$ at time j be v_j, and that obtained by using the input signal $\{\tilde{x}_j\}$ be $\tilde{v}_j$, respectively. Then,

$$v_j = \tilde{v}_j \qquad (j \geqq 0)$$

Proof. Let

$$\tilde{x}_j = (\tilde{x}_j, \tilde{x}_{j-1}, \cdots, \tilde{x}_{j-(n-1)})^t,$$
$$\tilde{y}_j = \langle w, \tilde{x}_j \rangle,$$
$$\tilde{X}_j = (\tilde{x}_j, \tilde{x}_{j-1}, \cdots, \tilde{x}_{j-(p-1)})^t,$$
$$\tilde{y}_j = (\tilde{y}_j, \tilde{y}_{j-1}, \cdots, \tilde{y}_{j-(p-1)})^t,$$

$\tilde{X}_j^+$: $\tilde{X}_j$ の Moore-Penrose generalized inverse matrix

Then, by the definition of APA, the coefficient vector $\tilde{v}_j$ for the input vector $\{\tilde{x}_j\}$ is successively determined as

$$\Delta \tilde{v}_j = \tilde{X}_j^+ (\tilde{y}_j - \tilde{X}_j \tilde{v}_j) \qquad (7)$$

$$\tilde{v}_{j+1} = \tilde{v}_j + \mu \Delta v_j \qquad (8)$$

Obviously, $\tilde{X}_j = a\, X_j$ and $\tilde{y}_j = a\, y_j$, and [8]

$$\tilde{X}_j^+ = \frac{1}{a} \tilde{X}_j$$

Consequently, from Eq. (7),

$$\Delta \tilde{v}_j = X_j^+ (y_j - X_j \tilde{v}_j)$$

Thus, if $\tilde{v}_j = v_j$, $\Delta\tilde{v}_j = \Delta v_j$, and $\tilde{v}_{j+1} = v_{j+1}$ follows from Eq. (8). From this and Eq. (6), the result is obtained by mathematical induction. (End of proof.)

From Property 3, it is seen that the behavior of the coefficient vector and the convergence speed in APA do not change if a nonzero constant is multiplied with the input signal. This implies that one does not have to consider the adjustment of the amplitude of the input signal, which is a very desirable property from the practical viewpoint. It was already described in Sect. 2 that the learning identification has the Properties 1 and 3, but those properties are shared in common by the whole APA.

4.3 The case where external disturbance exists

So far, it is assumed that no signal is impressed on the system of Fig. 1 other than the input signal $\{x_j\}$. In practical applications, however, there are many cases where other signals exist, which is not negligible. For example, in the application of the adaptive filter to the elimination of two input signals as in Fig. 6 [1], the signal from the noise source corresponds to $\{x_j\}$ in Fig. 1, which is required to identify the system to be identified (i.e., the transfer function from the noise source to the input terminal) and the signal from the signal source is not the one considered up to the point. The signal from the signal source, however, is the signal to be picked up, and if it is neglected, the whole problem will become meaningless.

The situation in Fig. 6 can be modelled as in Fig. 7. In the following, this system is used to consider the effect of the external disturbance on the coefficient vector v_j and the output signal $\{e_j\}$. In the model in Fig. 7, the following three coefficient vectors are defined.

(1) Let $\{y_j\}$ be the desired output. Let the coefficient vector, which is obtained

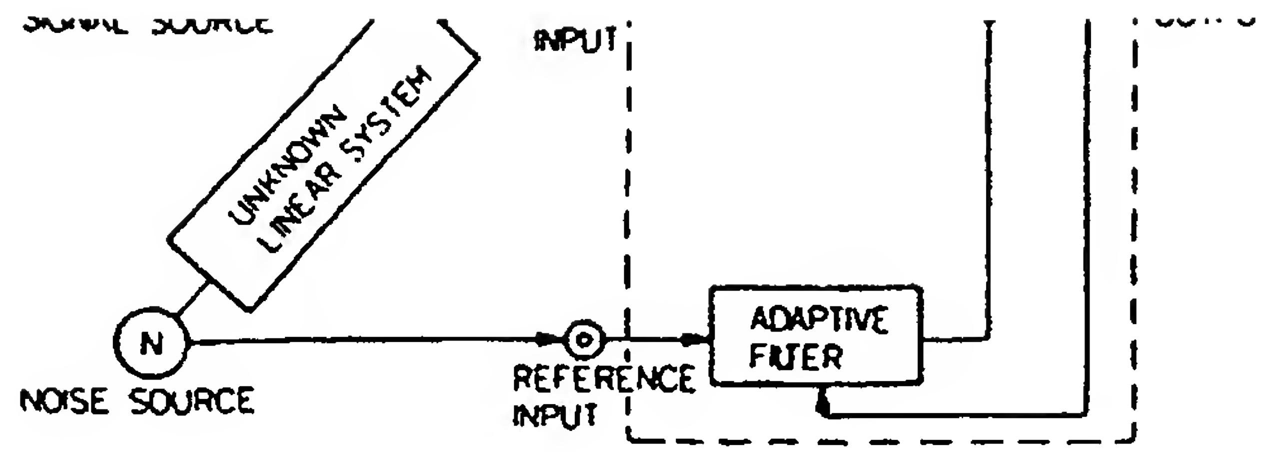

Fig. 6. Schematic diagram of a 2-input noise canceller using an adaptive filter.

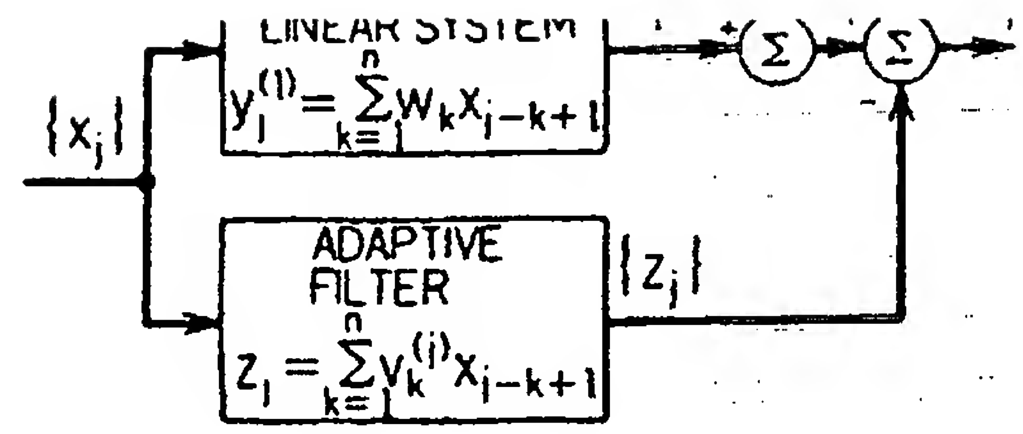

Fig. 7. System identification when a distu ance signal exists.

by successive modification by APA of the coefficient vector with the initial value v_0 be v_j.

(2) Let $\{y_j^{(1)}\}$ be the desired output. Let the coefficient vector which is obtained by successive modification by APA of the coefficient vector with the initial value v_0 be $v_j^{(1)}$.

(3) Let $\{y_j^{(2)}\}$ be the desired output. Then let the coefficient vector which is obtained by successive modification by APA of the coefficient vector with the initial value 0 be $v_j^{(2)}$.

Then, as can easily be verified

$$\boldsymbol{v}_j=\boldsymbol{v}_j^{(1)}+\boldsymbol{v}_j^{(2)} \tag{9}$$

The definition for $v_j^{(1)}$ is nothing but that for the coefficient vector without external disturbance, and $v_j^{(2)}$ is the term newly produced by the external disturbance.

Using Eq. (9), the output signal e_j can be decomposed as

$$e_j=e_j^{(1)}+e_j^{(2)}+e_j^{(3)}$$

where

$$e_j^{(1)}=y_j^{(1)}-\langle \boldsymbol{v}_j^{(1)},\ \boldsymbol{x}_j\rangle$$
$$e_j^{(2)}=y_j^{(2)}$$
$$e_j^{(3)}=-\langle \boldsymbol{v}_j^{(2)},\ \boldsymbol{x}_j\rangle$$

The signal $e_j^{(1)}$ is equal to the output signal without external disturbance. The signal $e_j^{(2)}$ is the applied disturbance itself; $e_j^{(3)}$ is the term produced by application of the external disturbance.

Consider a case where the output signal $\{e_j\}$ converges to 0 if there is no external disturbance. Then after a sufficient elapse of time.

$$e_j \fallingdotseq e_j^{(2)}+e_j^{(3)}$$

Thus consider how $e_j^{(3)}$ changes with th change of amplitudes of the input signal {: and the external disturbance $\{y_j^{(2)}\}$. Fro the definition of $v_j^{(2)}$,

$$\boldsymbol{v}_{j+1}^{(2)}=(I-\mu X_j^{+}X_j)\,\boldsymbol{v}_j^{(2)}+\mu X_j^{+}\boldsymbol{y}_j^{(2)}$$
$$\boldsymbol{y}_j^{(2)}=(y_j^{(2)},\ y_{j-1}^{(2)},\cdots,\ y_{j-(p-1)}^{(2)})^t$$

Using the notations

$$\Phi(j\ ;\ j)=I$$
$$\Phi(j+1\ ;\ j)=I-\mu X_j^{+}X_j$$
$$\Phi(k\ ;\ j)=\Phi(k\ ;\ k-1)\circ\Phi(k-1\ ;\ k-\dots$$
$$\circ\cdots\circ\Phi(j+1\ ;\ j)\ (k>j)$$
$$\boldsymbol{u}_j^{(2)}=\mu X_j^{+}\boldsymbol{y}_j^{(2)}$$

$v_j^{(2)}$ can be represented as

$$\boldsymbol{v}_j^{(2)}=\sum_{k=0}^{j-1}\Phi(j\ ;\ k)\,\boldsymbol{u}_k^{(2)} \qquad ($$

When $\{y_j^{(2)}\}$ is multiplied by a, Φ(j does not change and $u_k^{(2)}$ is multiplied by Consequently, by Eq. (10), $v_j^{(2)}$ is multip by a, and $e_j^{(3)}$ is multiplied by a. When is multiplied by b ($b \neq 0$), X_j is multipli by b and X_j^{+} is multiplied by 1/b, as is d scribed in the proof of Property 3. Consequently, Φ(k; j) does not change and u_k is multiplied by 1/b. By Eq. (10), $v_j^{(2)}$ is multiplied by 1/b, but $\boldsymbol{x}_j$ is multiplied by and $e_j^{(3)}$ does not change.

Consider the case where μ is changed When μ approaches 0, u_k approaches 0 and Φ k) approaches I. Consequently, by Eq. (10 $v_j^{(2)}$ approaches 0 amd $e_j^{(3)}$ approaches 0. In the application of the adaptive filter the noise elimination in two-input case, $\{e_j^{(2)}\}$ (= $\{y_j^{(2)}\}$) is the desirable signa and $\{e_j^{(3)}\}$ is the output noise. Consequently, the following property is derived from the above discussion.

and μ on $\{e_j(1)\} \sim \{e_j(5)\}$

Output component \	External disturbance $\{y_j(2)\}$ multiplied by a	Input signal $\{x_j\}$ multiplied by b	Relaxation coefficient $\mu \to 0$
$\{e_j(1)\}$	No change	Multiplied by b	Convergence speed $\to 0$
$\{e_j(2)\}$ $(= \{y_j(2)\})$	Multiplied by a	No change	No change
$\{e_j(3)\}$	Multiplied by a	No change	$\to 0$
$\{e_j(4)\}$	No change	Multiplied by b	No change
$\{e_j(5)\}$	No change	Multiplied by b	$\to 0$

Property 4. When APA is used in the noise elimination in two-input system, the signal-to-noise ratio of the output is independent of the input signal-to-noise ratio, and approaches infinity when the relaxation constant approaches 0.

4.4 Adaptive filter of insufficient order

So far, only the case is considered where the order of the system to be identified is the same as that of the adaptive filter. In practice, however, they usually do not coincide. No problem occurs when the order of the adaptive filter is larger than that of the system to be identified. Consider how the output signal $\{e_j\}$ changes when the order of the adaptive filter is not sufficient.

Assume that the system to be identified in Fig. 7 has an infinite order, and let the coefficients be $w_1, w_2, \cdots$. The output of this system at time j is given by

$$\sum_{k=1}^{\infty} w_k x_{j-k+1}$$

which is decomposed into two components as

$$y_j^{(1)} = \sum_{k=1}^{n} w_k x_{j-k+1}$$

$$y_j^{(3)} = \sum_{k=n+1}^{\infty} w_k x_{j-k+1}$$

As in the previous section, let $y_j(3)$ be the desired ouptut. Let the coefficient vector obtained by successive modification by APA of the coefficient vector with the initial value 0 be $v_j(3)$. Then v_j is decomposed into three components as

$$v_j = v_j^{(1)} + v_j^{(2)} + v_j^{(3)} \tag{11}$$

where v_j, $v_j(1)$ and $v_j(2)$ are vectors defined in the previous section. Using Eq. (11), e_j is decomposed into five terms as follows:

$$e_j = e_j^{(1)} + e_j^{(2)} + e_j^{(3)} + e_j^{(4)} + e_j^{(5)}$$

where $e_j(1)$, $e_j(2)$ and $e_j(3)$ are the variables defined in the previous section, and

$$e_j^{(4)} = y_j^{(3)}$$
$$e_j^{(5)} = -\langle v_j^{(3)},\ x_j \rangle$$

Consider how those terms change when the amplitude of $\{x_j\}$ is changed. When $\{x_j\}$ is multiplied by b $(b \neq 0)$, obviously $e_j(4)$ is multiplied by b. Letting

$$u_j^{(3)} = \mu X_j^{+} y_j^{(3)}$$
$$y_j^{(3)} = (y_j^{(3)},\ y_{j-1}^{(3)}, \cdots, y_{j-(p-1)}^{(3)})^t$$

we can write

$$v_j^{(3)} = \sum_{k=0}^{j-1} \Phi(j\ ;\ k)\, u_k^{(3)} \tag{12}$$

in the same way as for $v_j(2)$ in the previous section. When $\{x_j\}$ is multiplied by b $(b \neq 0)$, X_j^+ is multiplied by 1/b and $y_j(3)$ is multiplied by b. Consequently, $u_j(3)$ does not change and, by Eq. (12), $v_j(3)$ does not change. Thus, $e_j(5)$ is multiplied by b. In other words, both $e_j(4)$ and $e_j(5)$ are terms proportional to the amplitude of the input signal $\{x_j\}$.

Thus, when the order of the adaptive filter is not sufficient, a noise component proportional to the amplitude of the input signal $\{x_j\}$ appears in the output, and Property 4 does not apply. The effects of the input signal $\{x_j\}$, external disturbance $\{y_j(2)\}$ and the change of the value of μ on $\{e_j(1) \sim e_j(5)\}$ are summarized in Table 1.

5. Conclusions

The qualitative properties of the adaptive algorithm APA proposed in this paper are clarified by this study. However, some quantitative properties require further study. In the case where the convergence speed of the adaptive filter degrades greatly depending on the property of the input signal, an approach by lattice filter is also proposed [9]. A problem left for further study is to compare the merits and demerits of the lattice filter with APA.

A problem in APA is that the computational complexity for a modification increases with the order. Taking the recent hardware progress into consideration, the computational complexity will be overcome in the near future. APA with second or higher order will become practical in the future.

Acknowledgements. The authors acknowledge the suggestions and useful discussions made by Dr. T. Yasuhiro, Senior Researcher, and Dr. S. Yamamoto, Director, Acoust. and Audio Res. Div.

REFERENCES

1. B. Widrow et al. Adaptive Noise Cancelling: Principles and Applications, Proc. I.E.E.E., 63, 12 (Dec. 1975).
2. J. Nagumo and A. Noda. A Learning Method for System Identification, I.E.E.E. Trans. Automatic Control, AC-12, 3 (June 1967).
3. T. Araseki, Y. Ito and K. Ochiai. Microprocessor implementation of echo canceller, Tech. Rep., I.E.C.E., Japan, CS77-11 (Oct. 1977).
4. K. Ozeki and T. Umeda. A method of improving the convergence speed of adaptive filter, Tech. Rep. Speech Acoust. Soc. Jap., S-81-10 (May 1981).
5. K. Ozeki and T. Umeda. APA—Adaptive algorithm using affine projection, Nat. Conv., I.E.C.E., Japan, S1-23 (1982).
6. K. Ozeki and T. Umeda. Properties of the adapation algorithm based on the affine projection, Tech. Rep. Speech, S-82-53 (Nov. 1981).
7. K. Tanabe. Generalized inverse matrix (1) to (3), Oper. Res.. (Apr./June 1976)
8. S. Furuya. Generalized inverse matrix I, II, Math. Sci. (May/June 1967).
9. L.J. Griffiths. An adaptive lattice structure for noise-cancelling applications, Proc. ICASSP, pp. 87-90 (April 1978).